

基于总体局部均值分解方法的心律失常特征提取与分类

陈敏,王尧芬

上海工程技术大学电子电气工程学院,上海 201620

【摘要】针对心电信号自动分类技术中的特征提取,提出一种新的特征提取方法——总体局部均值分解(ELMD)方法。该方法首先对心电信号加入不同的高斯白噪声,然后进行局部均值分解得到若干乘积函数(PF)分量,求取多次分解后的PF分量均值。多次加入噪声及分量平均的过程可以克服基本局部均值分解方法存在的模态混叠问题。选取较优的前4个PF分量进行特征计算,将得到的特征向量矩阵送入支持向量机对正常心电信号和4种常见的心律失常信号进行分类。从MIT-BIH心律失常数据库的分类结果来看,ELMD总体分类准确率达到99.61%,高于一般方法,证明了ELMD方法的有效性。

【关键词】心电信号;总体局部均值分解;特征提取;心律失常分类;支持向量机

【中图分类号】R318;TP391

【文献标志码】A

【文章编号】1005-202X(2019)10-1211-06

Feature extraction and classification of arrhythmia using ensemble local mean decomposition method

CHEN Min, WANG Raofen

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

Abstract: Aiming at feature extraction in electrocardiogram (ECG) automatic classification technology, a new method of feature extraction, ensemble local mean decomposition, is proposed. Firstly, different Gaussian white noises are added to ECG signals. Then, several product function (PF) components are obtained by local mean decomposition, and the average value after multiple decompositions is calculated as the final PF components. This method can solve the mode aliasing problem of local mean decomposition by multiple additions of noise and component averaging. The top 4 features of PF components are selected for feature calculation. The obtained feature vector matrix is used in support vector machine to classify normal ECG signals and 4 common kinds of arrhythmia ECG signals. Finally, the verification with MIT-BIH arrhythmia database shows that the total classification accuracy of the proposed method reaches 99.61%, higher than conventional methods, which confirms the effectiveness of ensemble local mean decomposition method.

Keywords: electrocardiogram signal; ensemble local mean decomposition; feature extraction; arrhythmia classification; support vector machine

前言

在社会老龄化和城市化进程加快、居民不健康生活方式盛行的背景下,中国心血管病患率和死亡率一直处于不断上升阶段,并呈现出低龄化趋势

且在低收入群体中快速增长及个体聚集现象^[1]。心电图是人体心电信号的反映,心脏专家通过对心电图的观察,可以识别心电信号的异常,从而判断人体心脏的健康状况。除医院内广泛使用的Holter心电监护仪外,家用可穿戴心电设备也逐渐普及^[2-3]。这就需要对心电信号进行自动分类,并且有很好的实时性,从而实现对心脏健康状况的实时监测,及时发现心律失常并采取相应措施^[4]。

整个心电信号自动分类过程中,特征提取是至关重要的部分,众多学者对此进行了研究。Li等^[5]使用小波包和近似熵对心电信号进行特征提取,小波分析方法具有多分辨率分析的特点,在时频域能有

【收稿日期】2019-06-20

【基金项目】国家自然科学基金(61803255, 71701124);上海市自然科学基金(18ZR1416700)

【作者简介】陈敏,在校研究生,研究方向:生理信号处理和模式识别, E-mail: 1723346394@qq.com

【通信作者】王尧芬,博士,副教授,研究方向:智能建模,生理信号分析及疲劳识别, E-mail: rfwangsues@163.com

效表征信号的局部特征,但存在小波基函数的选择等问题;Rajesh等^[6]使用经验模态分解(Empirical Mode Decomposition, EMD)方法对心电信号进行分解,对得到的IMF分量进行熵等特征的计算得到特征向量。EMD是一种自适应的信号处理方法,自适应分解得到的若干个IMF分量在时频上都有其物理意义。然而EMD分解中会出现端点效应、模态混叠、过包络和欠包络等问题^[7],针对这些缺点出现过很多改进的方法,比如总体经验模态分解^[8]、总体完备经验模态分解^[9-10]等,但是这些改进方法依然不能够有效解决EMD存在的问题。胡健等^[11]提出使用新的特征提取方法——局部均值分解(Local Mean Decomposition, LMD)方法对心电信号进行自适应分解,该方法能够解决EMD方法的端点效应、过包络、欠包络问题,但仍然存在模态混叠问题。因此本文对LMD方法进行改进,提出一种总体局部均值分解(Ensemble Local Mean Decomposition, ELMD)方法,并应用ELMD方法对心电信号进行分析。ELMD方法通过对预处理后的心电信号多次加入不同的高斯白噪声后再进行局部均值分解,对每次得到的若干个乘积函数分量(Product Function, PF)求均值,得到最后的PF分量。该方法分解得到的PF分量能够克服由信号间歇性引起的信号时频分布的混叠,即在一个PF分量中只出现信号相同时间尺度的特征。

对优化后的PF分量进行特征计算,得到一组能有效代表心电信号特性的特征向量。最后送入支持向量机进行分类,实现对正常心电信号和4种常见心律失常心电信号的自动分类。

1 特征提取方法

1.1 LMD方法

LMD是一种新的自适应时频分析方法,它能够非线性、非平稳的复杂信号自适应地分解为若干个具有物理意义的PF分量和一个残余分量,在机械故障诊断和脑电信号分析中得到广泛应用^[12-14]。具体算法如下:找出原始信号 $x(t)$ 所有的局部极大值点和极小值点 n_j ,求取相邻两个极值点的均值,用滑动平均法进行平滑处理,得到局部均值函数 $m_{11}(t)$,并通过局部均值点计算包络估计值 a_j :

$$a_j = |n_j - n_{j+1}|/2 \quad (1)$$

将 a_j 用直线连接起来,采用滑动平均法对折线进行平滑处理,得到包络估计函数 $a_{11}(t)$;从 $x(t)$ 中分离出局部均值函数,得到 $h_{11}(t)$;将分离得到的 $h_{11}(t)$ 除以 $a_{11}(t)$ 进行解调,得到调频信号:

$$s_{11}(t) = h_{11}(t)/a_{11}(t) \quad (2)$$

重复上述步骤 n 次,直到 $s_{1n}(t)$ 为纯调频函数,此时:

$$h_{1n}(t) = s_{1(n-1)}(t) - m_{1n}(t) \quad (3)$$

$$s_{1n}(t) = h_{1n}(t)/a_{1n}(t) \quad (4)$$

则 $x(t)$ 分解出的第一个PF分量为包络信号 $a_1(t)$ 和纯调频信号 $s_{1n}(t)$ 的乘积,即:

$$PF_1 = a_1(t)s_{1n}(t) \quad (5)$$

从 $x(t)$ 中分离出 $PF_1(t)$,将得到的新信号 $u_1(t)$ 作为源信号迭代上述步骤 k 次,直到 $u_k(t)$ 为单调函数。最后, $x^i(t)$ 被分解为 k 个PF分量和一个残余分量 $u_k(t)$,表示为:

$$x(t) = \sum_{i=1}^k PF_i(t) + u_k(t) \quad (6)$$

1.2 ELMD方法

ELMD方法是对LMD方法的一种改进。ELMD的本质是对信号加入不同的高斯白噪声后进行局部均值分解,对每次得到的若干个PF分量求均值作为最后的结果。其具体的计算流程如下:

(1)对去噪之后的心电信号 $x(t)$ 加入不同的高斯白噪声:

$$x^i(t) = x(t) + \omega^i(t) \quad (7)$$

其中, $\omega^i(t)(i=1, 2, \dots, M)$ 是不同的均值为0、方差为 ε 的高斯白噪声。

(2)对 $x^i(t)$ 进行LMD分解,得到 k 个PF分量和一个残余分量 $u_k(t)$,表示为:

$$x^j(t) = \sum_{i=1}^k PF_i(t) + u_k(t) \quad (8)$$

(3)计算 M 次分解得到的PF分量 p 和余量 r 的均值:

$$\begin{cases} \bar{p}_i = \sum_{i=1}^M p_{ik}/M \\ \bar{r} = \sum_{i=1}^M r/M_i \end{cases} \quad (9)$$

其中, \bar{p}_i 、 \bar{r} 即为ELMD分解的结果。

心电信号经过ELMD的分解,原信号的特征信息会被分布到不同的时间特征尺度上,从而放大了原信号隐含的特征信息。

2 心律失常特征提取及分类

2.1 数据来源

本文数据采用美国麻省理工学院心律失常数据库MIT-BIH。MIT-BIH数据库广泛应用于学术研究和工程实施^[15-16],它由48段长度为半小时的记录组成,来自47个个体。每个心拍都至少由两名心脏学

专家进行心律失常类型的标注,包括束支传导阻滞、房性早搏、室性早搏、房颤、起搏心跳、ST段改变等15种心律异常。本文对4种常见心律失常疾病:左束支传导阻滞(Left Bundle Branch Block, LBBB)、右束支传导阻滞(Right Bundle Branch Block, RBBB)、房性早搏(Atrial Premature Contraction, APC)、室性早搏(Premature Ventricular Contraction, PVC)和正常(Normal, N)心电信号进行分类,选择的数据编号及心拍样本数量如表1所示。

表1 心电数据编号及样本数量

Tab.1 Electrocardiogram (ECG) data number and sample size

类别	样本数量	数据编号
N	800	100, 103, 105, 113
LBBB	800	109, 111, 207, 214
RBBB	800	118, 212, 124, 231
APC	800	232, 209, 220, 118, 222
PVC	800	200, 119, 233, 118

N:正常信号;LBBB:左束支传导阻滞;RBBB:右束支传导阻滞;
APC:房性早搏;PVC:室性早搏

2.2 心电信号去噪及心拍分割

心电信号在采集过程中,必然会受到各种类型噪声的干扰,如工频干扰、基线漂移和高频干扰等^[17]。本文使用Sing等^[18]提出的方法,选择db6小波基函数将心电信号分解为9层。由于第9层近似系数的频率范围为0~0.351 Hz,主要是基线漂移,因此不用于重构。此外,90~180和45~90的频带范围包含工频干扰和高频干扰,也不用于重构。通过对第3层到第9层的细节系数重构,实现对心电信号的去噪。

去噪后使用WFDB工具箱的ecgpuwave函数,实现对QRS波群的检测。如图1所示,实现了对记录编号100的心电信号QRS波群的检测。在QRS波群位置前取120个点,在QRS波群位置后取180个点进行心拍分割。分割后的单个心拍如图2所示。

2.3 基于ELMD方法的心电信号分析和特征提取

对预处理后的单个心拍使用ELMD方法进行自适应分解。其中,设定高斯白噪声的方差 ε 值为0.2,加入白噪声的次数 M 值为100。利用白噪声信号具有频率分布均匀的统计特性,且白噪声的均值为0,不同的白噪声加入心电信号中形成混合信号后,进行多次分解得到的多个PF分量平均值能够自动剔除白噪声成分。在对白噪声进行分解时,ELMD频域上表现为一个高通滤波器和一系列连续的带通滤波器组的组合。混合信

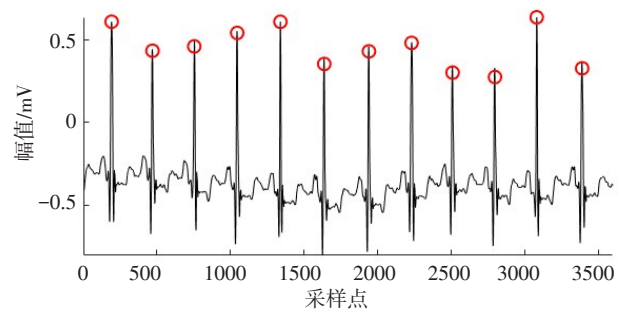


图1 QRS波群检测

Fig.1 QRS complex detection

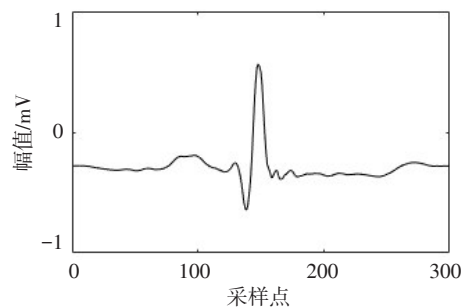


图2 预处理后的单个心拍

Fig.2 Single heartbeat after preprocessing

号经ELMD分解后,心电信号中的所有尺度会自动地分解到白噪声所确定的滤波器组中与之相关的通频带中,因此能够显著减少模态混叠现象。

对5种不同类型的心拍进行ELMD分解,均能得到4个PF分量和一个残余分量。图3是5种不同类型心电信号的ELMD分解结果,分别选取记录编号为100、109、118、200和232的心电数据。可以观察到每种心律失常类型所对应的PF分量波形各不相同,由于残余分量所包含的信息较少,所以选择前4个PF分量进行特征提取。

2.4 心电信号特征提取

ELMD分解后,计算每个PF分量的样本熵(SE)、平均功率(MP)、奇异值(SV)和标准差(SD),得到16维特征。其中,SE反映信号的复杂程度^[19];MP反映每个PF分量的频带信息;SV反映信号的稳定性^[20];SD反映信号的变化特性。组成的特征矩阵如表2所示。

每类心电信号类型选择800个样本进行分析,采用卡方检验方法分析每一个特征与所属类别的相关性并进行评分,对分数求log后,得到图4的相关系数图。可以看出每一个特征对心电信号的分类都起到一定的作用。图5为随机选择的3个特征的箱型图表示,可以观察到每类心电信号的特征值都有明显的差异分布,即每类心电信号的特征值都具有较好的分类辨识度。

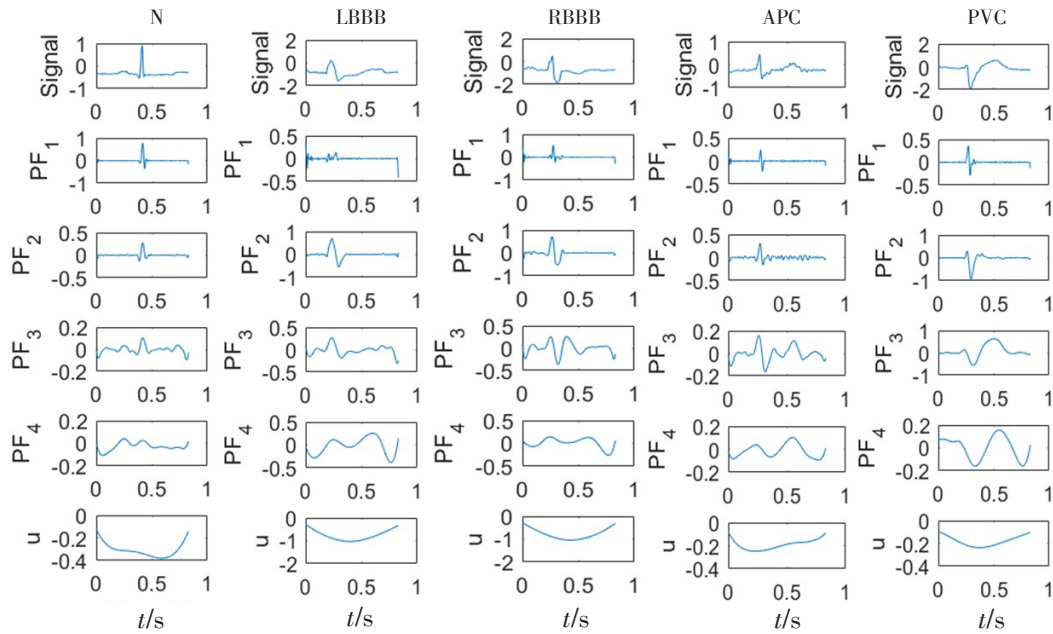


图3 N,LBBB,RBBB,APC,PVC的ELMD分解图

Fig.3 Ensemble local mean decomposition (ELMD) diagrams of N, LBBB, RBBB, APC and PVC

表2 心电特征矩阵名称

Tab.2 Names of ECG feature matrixes

特征	PF1	PF2	PF3	PF4
样本熵	SE(PF1)	SE(PF2)	SE(PF3)	SE(PF4)
平均功率	MP(PF1)	MP(PF2)	MP(PF3)	MP(PF4)
奇异值	SV(PF1)	SV(PF2)	SV(PF3)	SV(PF4)
标准差	SD(PF1)	SD(PF2)	SD(PF3)	SD(PF4)

2.5 心电信号分类

首先对数据集进行划分,在每类心电信号中随机选取 70%的数据作为训练集,剩余的 30%作为测试集,得到 2 800 组训练样本和 1 200 组测试样本,分类模型选择支持向量机(SVM)。SVM通过引入高斯核函数,将输入向量映射到一个高维的特征向量空间中,并在该特征空间中构造最优分类超平面,实现对正常信号和 4 种心律失常类型心电信号的分类。其中,使用遗传算法(GA)和十折交叉验证对 SVM 中涉及到的惩罚因子 C 和核函数参数 η 进行寻优操作。最后得到最优惩罚因子的值为 22.633 4,最优核函数参数的值为 3.049 4。使用参数优化后的 SVM 对心电信号进行分类。

3 结果与分析

本文采用总准确率(Total Accuracy, TA)、灵敏度(Sensitivity, Se)和特异性(Specificity, Sp)3 种方法对

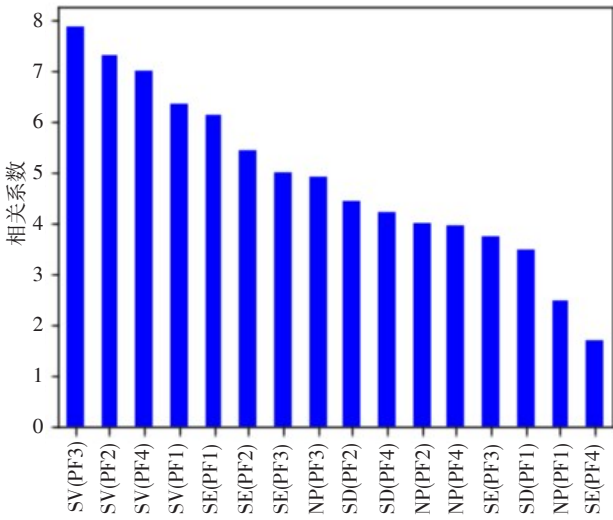


图4 心电特征与心律失常的相关性

Fig.4 Correlation between ECG features and arrhythmia

心律失常分类效果进行分析和评估。公式如下:

$$TA = \frac{TP + TN}{TP + FP + FN + TN} \times 100\% \tag{10}$$
$$Se = \frac{TP}{TP + FN} \times 100\% \tag{11}$$
$$Sp = \frac{TN}{TN + FP} \times 100\% \tag{12}$$

其中,TP 表示当前心拍类型正确分类的心拍个数;FP 表示不属于当前心拍类型被错分为当前心拍类型的心拍个数;TN 表示属于当前心拍类型被错分为其他类型的心拍个数;FN 表示不属于当前心拍类型且被错分为其他类型的心拍个数。

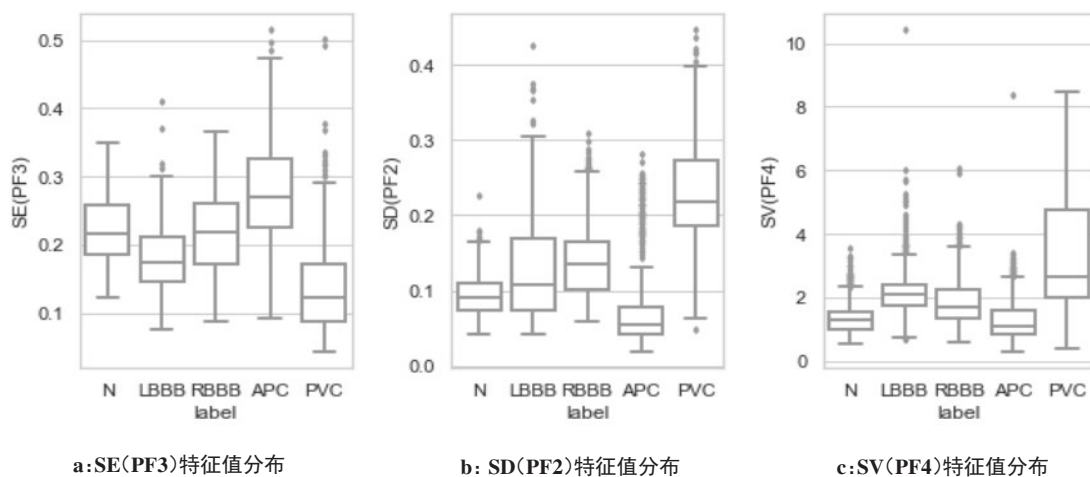


图5 心电特征值分布

Fig.5 ECG feature value distribution

将得到的特征矩阵输入参数优化后的SVM进行分类,得到的混淆矩阵如表3所示。通过表3可以观察到,使用ELMD方法进行特征提取的分类结果中,5种心电信号类型都达到了近乎完全正确的分类,有4个RBBB错分为APC,相对来说,这两种心拍类型容易混淆。通过对混淆矩阵的计算,得到表4的评价指标结果。可以看到除RBBB的灵敏度外,其他评价指标均达到99%以上,达到了很好的分类效果。

表3 分类结果的混淆矩阵表示(ELMD方法)

Tab.3 Confusion matrix of classification results (ELMD method)

类别	N	LBBB	RBBB	APC	PVC
N	238	0	2	0	0
LBBB	0	238	0	0	2
RBBB	1	0	235	4	0
APC	0	0	1	229	0
PVC	0	1	0	0	239

表4 心律失常分类的实验结果(ELMD方法)

Tab.4 Results of classification of arrhythmia (ELMD method)

类别	总准确率/%	灵敏度/%	特异性/%
N		99.17	99.90
LBBB		99.17	99.90
RBBB	99.61	97.92	99.69
APC		99.58	99.45
PVC		99.58	99.79

本文还对相同的心电信号样本使用LMD方法进行

自适应分解,对得到的3个PF分量进行相同的特征提取并使用相同的分类器进行分类。得到的混淆矩阵和评价指标结果如表5和表6所示。通过表3和表5对比可以看出,LMD方法RBBB和APC的区分表现得更糟糕,RBBB有12个样本被错分为APC,有9个APC被分类为RBBB,且其他类型分类错误的样本数量均高于表3中的数量。在表6中,LMD方法LBBB、RBBB和APC的灵敏度均低于97%,尤其是RBBB的灵敏度仅为91.67%。灵敏度越低表示漏诊率越高,说明很多RBBB的心拍错分为其他类型。验证了ELMD方法相比于LMD方法具有明显的优越性。

除与LMD方法进行对比外,本文将使用ELMD方法得到的结果与其他特征提取方法的分类结果进行比较,分类器均为SVM且分类种类相同。如表7所示,使用EMD^[6]和小波变换(WT)、近似熵(ApEn)^[5]作为特征提取方法的分类精确度均比本文方法低。因此证明本文的实验方法是一种效果很好的心电信号自动识别方法。

4 结论

有效的心电特征提取对心律失常的正确分类有着重要意义。本文使用了一种新的心电信号特征提取技术—ELMD方法对预处理后的心电信号进行特征提取,并使用遗传算法优化后的SVM进行分类。通过MIT-BIH数据库的4 000个样本进行分析后,评价指标结果表明使用ELMD方法提取的特征比使用LMD方法提取的特征更具代表性,且准确率达到了99.61%,结果优于其他文献,验证了本文方法的可靠性。本次研究主要是在特征提取方法上进行了创新,没有考虑病人的差异性。因此,下一步的研究目标计划是让训练集和

表5 分类结果的混淆矩阵表示(LMD方法)

Tab.5 Confusion matrix of classification results (LMD method)

类别	Normal	LBBB	RBBB	APC	PVC
Normal	234	5	1	0	0
LBBB	4	229	4	0	3
RBBB	5	3	220	12	0
APC	1	2	9	228	0
PVC	0	0	0	0	240

表7 与其他文献的结果比较

Tab.7 Comparison with the results from other literatures

作者	特征提取方法	总准确率/%
Rajesh等 ^[6]	EMD	98.72
Li等 ^[5]	WT和ApEn	97.78
本文	ELMD	99.61

测试集的样本来自不同的病人,找到有效的特征提取方法实现心电信号的精确分类,从而得到更好的适用性。

【参考文献】

- [1] 马丽媛,吴亚哲,王文,等.《中国心血管疾病报告2017》要点解读[J]. 中国心血管杂志, 2018, 23(1): 3-6.
MA L Y, WU Y Z, WANG W, et al. Interpretation of the report on cardiovascular diseases in China (2017) [J]. Chinese Journal of Cardiovascular Medicine, 2018, 23(1): 3-6.
- [2] 吴敏,谢云,郭洋. 便携式心电监测系统的硬件设计与实现[J]. 中国医学物理学杂志, 2018, 35(2): 210-213.
WU M, XIE Y, WU Y. Hardware design and implementation of portable electrocardiogram monitoring system [J]. Chinese Journal of Medical Physics, 2018, 35(2): 210-213.
- [3] XIA Y F, XIE Y Q. A novel wearable electrocardiogram classification system using convolutional neural networks and active learning [J]. IEEE Access, 2019(7): 7989-8001.
- [4] YILDIRIM Ö, PLAWEK P, TAN R S, et al. Arrhythmia detection using deep convolutional neural network with long duration ECG signals [J]. Comput Biol Med, 2018, 102: 411-420.
- [5] LI H Q, YUAN D Y, MA X D, et al. A new ECG signal classification based on WPD and ApEn feature extraction [J]. Circuits Syst Signal Process, 2016, 35(1): 339-352.
- [6] RAJESH K N, DHULI R. Classification of ECG heartbeats using nonlinear decomposition methods and support vector machine [J]. Comput Biol Med, 2017, 87: 271-284.
- [7] 程军圣,张亢,杨宇,等. 局部均值分解与经验模态分解的对比研究[J]. 振动与冲击, 2009, 28(5): 13-16.
CHENG J S, ZHANG K, YANG Y, et al. Comparison between the methods of local mean decomposition and empirical mode decomposition [J]. Journal of Vibration and Shock, 2009, 28(5): 13-16.
- [8] CUI X R, CHANG H C, LIN L Y, et al. Prediction of atrial fibrillation

表6 心律失常分类的实验结果(LMD方法)

Tab.6 Results of classification of arrhythmia (LMD method)

类别	总准确率/%	灵敏度/%	特异性/%
Normal		97.50	98.96
LBBB		95.42	98.96
RBBB	98.28	91.67	98.54
APC		95.00	98.33
PVC		100.00	99.69

recurrence before catheter ablation using an adaptive nonlinear and non-stationary surface ECG analysis [J]. Physica A, 2019(514): 9-19.

- [9] RAJESH K, DHULI R. Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier [J]. Biomed Signal Process Control, 2018, 41: 242-254.
- [10] 张欢,刘燕,佟宝同,等. 联合改进CEEMD与近似熵的脑电信号降噪方法[J]. 计算机工程, 2016, 42(4): 313-317.
ZHANG H, LIU Y, TONG B T, et al. Electroencephalogram denoising method combining improved CEEMD and approximate entropy [J]. Computer Engineering, 2016, 42(4): 313-317.
- [11] 胡健,张凯. 基于LMD与样本熵的心电信号异常分类研究[J]. 计算机科学, 2018, 45(10A): 306-310.
HU J, ZHANG K. Classification of ECG signals based on LMD and sample entropy [J]. Computer Science, 2018, 45(10A): 306-310.
- [12] HAO R, JIANFENG Q, YI C, et al. An adaptive unsupervised classification of seizure method based on LMD-MSSE with EEG signals [J]. Biomed Res, 2017, 28(7): 3151-3158.
- [13] ZHANG T, CHEN W Z. LMD based features for the automatic seizure detection of EEG signals using SVM [J]. IEEE Trans Neural Syst Rehabil Eng, 2017, 25(8): 1100-1108.
- [14] 赵利民,朱晓军. 基于局部均值分解与样本熵的脑电信号特征提取与分类[J]. 计算机工程, 2017, 43(2): 299-303.
ZHAO L M, ZHU X J. Feature extraction and classification of electroencephalogram signals based on local mean decomposition and sample entropy [J]. Computer Engineering, 2017, 43(2): 299-303.
- [15] DILMAC S, DOKUR Z, ÖLMEZ T. Comparative analysis of MABC with KNN, SOM, and ACO algorithms for ECG heartbeat classification [J]. Turk J Electr Eng Co, 2018, 26(6): 2820-2831.
- [16] YILDIRIM Ö. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification [J]. Comput Biol Med, 2018, 96: 189-202.
- [17] MOURAD N. ECG denoising algorithm based on group sparsity and singular spectrum analysis [J]. Biomed Signal Process Control, 2019, 50: 62-71.
- [18] SING B, TIWARI A. Optimal selection of wavelet basis function applied to ECG signal denoising [J]. Digit Signal Process, 2006, 16(3): 275-287.
- [19] HORIE T, BURIOKA N, AMISAKI T, et al. Sample entropy in electrocardiogram during atrial fibrillation [J]. Yonago Acta Med, 2018, 61(1): 49-57.
- [20] WANG D, SI Y J, YANG W Y, et al. A novel heart rate robust method for short-term [J]. Appl Sci, 2019, 9(1): 1-8.

(编辑:薛泽玲)